

sequence of the human *h* gene

GAATTCGGTCGGCAGATATGCAGCAATTACCGGGCGGGTCCTCTCAGCCAGCGCGGGGAGGCAGCGCGCGCGCAGCACCGCGGAACCGCACCGAGGAGAGAGCCAGCCCCCGCC

CTCGCCCCCTTCGGTCCCCACCCCCATCCCGCGCGCCAGCAGGCTCCCGCGCTCTCCCTGTTTCTCCTCCTGCTGCGGCTGCGCTCTCGCACTCATCTCTC

CGCGGGCGCGTCCGCGCAGCTCCGTGCTCCCGCGCGCCAGCCTCTCCGGCGCGCGCTCCCTAAGGATGGTACTGATTTTCCCGCGCACAGGAGACCGGGCTGAGCGCGCGCGCGCGG

CTCGCCTCTCTCCGAGCACGCCAGCGCTGGGACGGC
Met Arg Thr Leu Ala Cys Leu Leu Leu Gly Cys Gly Tyr Leu Ala His Val Leu Ala
ATG AGG ACC TTG GCT TGC CTC CTG CTC CTC GGC TGC GGA TAC CTC GCC CAT GTT CTG GCC

30. Glu Glu Ala Glu Ile Pro Arg Glu Val Ile Glu Arg Leu Ala Arg Ser Gln Ile His Ser Ile Arg Asp Leu Gln Arg Leu Glu Ile
 GAG GAA GCC GAG ATC CCC CGC GAG GTG ATC GAG AGG CTC GCC CGC AGT CAG ATC CAC AGC ATC CGG GAC CTC CAG CGA CTC CTC GAG ATA

[illegible]

CCC ATT	CGG	ACC	AAC	AGA	ACC	ATC	GAG	GAA	GCT	GTC	CCC	GCT	GTC	TGC	AAG	ACC	AGG	ACG	GTC	ATT	TAC	GAG	ATT	CCT	CGG	ACT	CAG	GTC	
Pro	Ile	Arg	Arg	Lys	Arg	Ser	Ile	Glu	Glu	Ala	Val	Pro	Ala	Val	Cys	Lys	Thr	Arg	Thr	Val	Ile	Tyr	Glu	Ile	Pro	Arg	Ser	Gln	Val
90																												100	
110																												110	

	120	130	140
Asp Pro Thr Ser Ala Asn Phe Leu Ile Trp Pro Pro Cys Val Glu Val Lys Arg Cys Thr Gly Cys Cys Asn Thr Ser Ser Val Lys Cys			

GAC CCC ACG TCC GCC AAC TTC CTC ATC TGG CCG CCC GAG GTG AAA CGC TGC ACC AGC GGC ICC AAC ACQ AGC AGC AGT
170

160

Gln Pro Ser Arg Val His Arg Ser Val Lys Val Ala Lys Val Glu Tyr;Val Arg Lys Lys Pro Lys Leu Lys Glu Val Gln Val Arg
CAG CCC TCC CGC GTC CAC CAC CAC CCG AGC GTC AAG CTG GCC AAG GTG GAA TAC GTC AGC AAG CCA AAA TTA AAA GAA GTC CAG GTG AGC
150

180	190	200
Leu Glu Glu His Leu Glu Cys Ala Cys Ala Thr Thr Ser Leu Aan Pro Asp Tyr Arg Glu Glu Asp Thr Gly Arg Pro Arg Glu Ser Gly	TTA GAG GAG CAT TTC GAG TGC GCC TGC GCG ACC ACA AGC CTG AAT CCG GAT TAT CCG GAA GAG GAC ACG GGA AGG CCT AGG GAG TCA GGT	

210 211
Lys Lys Arg Lys Arg Lys Arg Leu Lys Pro Thr OC
AAA AAA CCG AAA AAA AGA AAA AGG TTA AAA CCC ACC TAA AGCAGCCAACCCAGATGTCAGGTGAGGATGAGCCGAGCCCTTCTCGCACATGCGATGTACATGCCGTGTTA

CAATTCCTGAACCTACTACTATCGGTGCTTTATTGCCACGTGCGGTCTTTGTTCTCCTCCCTGAAAACGTGCTCGAGAACACTCGGAGAACAAAGACAGACAGTCGCACATTGTTTAAAT

GTGACATCAAGCAAGTATTGTAGCCACTGCGTAGCAGTAAGANGCTTCTCTGTCAAAGAGAGAGAGAGAGAAAGCAAAAAAAGGAATTC

Translated Mol. Weight = 24046.60

Figure 1

GAATTCGGTCCGCAAAATATGCAGATTACCGGCCGGGTCTCCTGTAAGCCAGCCGGGGAGGCAGCGCGCGCGGCGGCGGAGCACCGGGAACCGCACCGAGGAAGAGCCAGCGCCCGCC

CTCGCCCCCTTCGGTCCCGACCCCACATCCCGCGCGCCAGGAGGCTCCCGCGCTCGCGGCACTCCCTGTTCTCCTCTCTGGCTGGGCTGCTCTCCGCACTCATCTGCTC

GCGGGCGCGCTCCGCCACGCTCCGTCTCCCGCGCCACCCTCTCCGGGCGCGCTCCCTAMGGGATGGTAGTATTTCCCGCCACAGGAGACCGGCTGGAGCGCGCCCGCCGCGG

Met Arg Thr Leu Ala Cys Leu Leu Leu Leu Leu Gly Cys Gly Tyr Leu Ala His Val Leu Ala
ATG AGG ACC TTG GCT TGC CTG CTC CTG CTC GGC TGC GGA TAC CTC GCC CAT GTT CTG GCC

C TGCGCTCTCTCCGAGCAGCCAGCGCTCGGGAGCGG

30 Met Arg Thr Leu Ala Cys Leu Leu Leu Leu Leu Gly Cys Gly Tyr Leu Ala His Val Leu Ala
ATG AGG ACC TTG GCT TGC CTG CTC CTG CTC GGC TGC GGA TAC CTC GCC CAT GTT CTG GCC

Glu Glu Ala Glu Ile Pro Arg Glu Val Ile Glu Arg Glu Ala Arg Ser Glu Ile His Ser Ile Arg Asp Leu Gln Arg Leu Leu Glu Ile
CAG GAA GCC GAG ATC CCC CGC GAG ATC GAG AGG CTG GCC CGC AGT CAG ATC CAC AGC ATC CGG CAC CTC CAG CGA CTC CTG GAG ATA

60 Asp Ser Val Gly Ser Glu Asp Ser Thr Ser Leu Arg Ala His Gly Val His Ala Thr Lys His Val Pro Glu Lys Arg Pro Leu
GAC TCC GTA GGG AGT GAG GAT TCT TTG GAC ACC ACC CTG AGA GCT CAC GGG GTC CAT GCC ACT AAG CAT GTG CCC GAG AAG CGG CCC CTG

90 Pr Ile Arg Arg Lys Arg Ser Ile Glu Glu Ala Val Pro Ala Val Cys Lys Thr Arg Thr Val Ile Tyr Glu Ile Pro Arg Ser Gln Val
CCC ATT CCG AGG AAG AGA ACC ATC GAG GAA GCT GTC CCC GCT GTC TGC ANG. ACC AGG AGC GTC ATT TAC GAG ATT CCT CGG AGT CAG GTC

120 Asp Pro Thr Ser Ale Asn Phe Leu Ile Trp Pro Pro Cys Val Glu Val Lys Arg Cys Thr Gly Cys Cys Asn Thr Ser Ser Val Lys Cys
GAC CCC ACG TCC GCC AAC TTC CTG ATC TGG CCC CCG TGC GTC GAG GTG AAA CGC TGC ACC GGC TGC TGC AAG ACC AGC AGT GTC AAG TGC

150 Gln Pro Ser Arg Val His His Arg Ser Val Lys Val Ala Lys Val Glu Tyr Val Arg Lys Lys Pro Lys Leu Lys Glu Val Gln Val Arg
CAG CCC TCC CGC GTC CAC CAC CGC AGC GTC AAC GTG GCC AAG GTG GAA TAC GTC AGG AAG AAG CCA AAA TTA AAA GAA GTC CAG GTG ACG

180 Leu Glu Glu His Leu Glu Cys Ala Thr Thr Ser Leu Asn Pro Asp Tyr Arg Glu Glu Asp Thr Asp Val Arg OP
TTA GAG GAG CAT TTG GAG TGC GCC TGC GCG ACC ACA AGC CTG AAT CCG GAT. TAT CCG GAA GAG GAC ACG GAT GTG AGG TGA GGATGAGCGCGCA

GCCCTTCTCGGACATGGATGATACATGGCGTGTTACATTCTCTGAACCTACTATGTACGGTGCTTTATTGCCAGTGTGGGTCTTCTCTCCGTGAAAACCTGTCTCGAGAACAC

TCCGGAGAACAAAGACAGTCACATTTCTTTAATGTGACATCAAGCAAGTATTGTAGCACTCGGTGAGCAGTAAAGAAGCTTCTTCTCAAAGACAGACAGAAAGAAAAA

AAGCAATTG

AACGAATTC

Translated Mol. Weight = 22256.17

Figure 2.

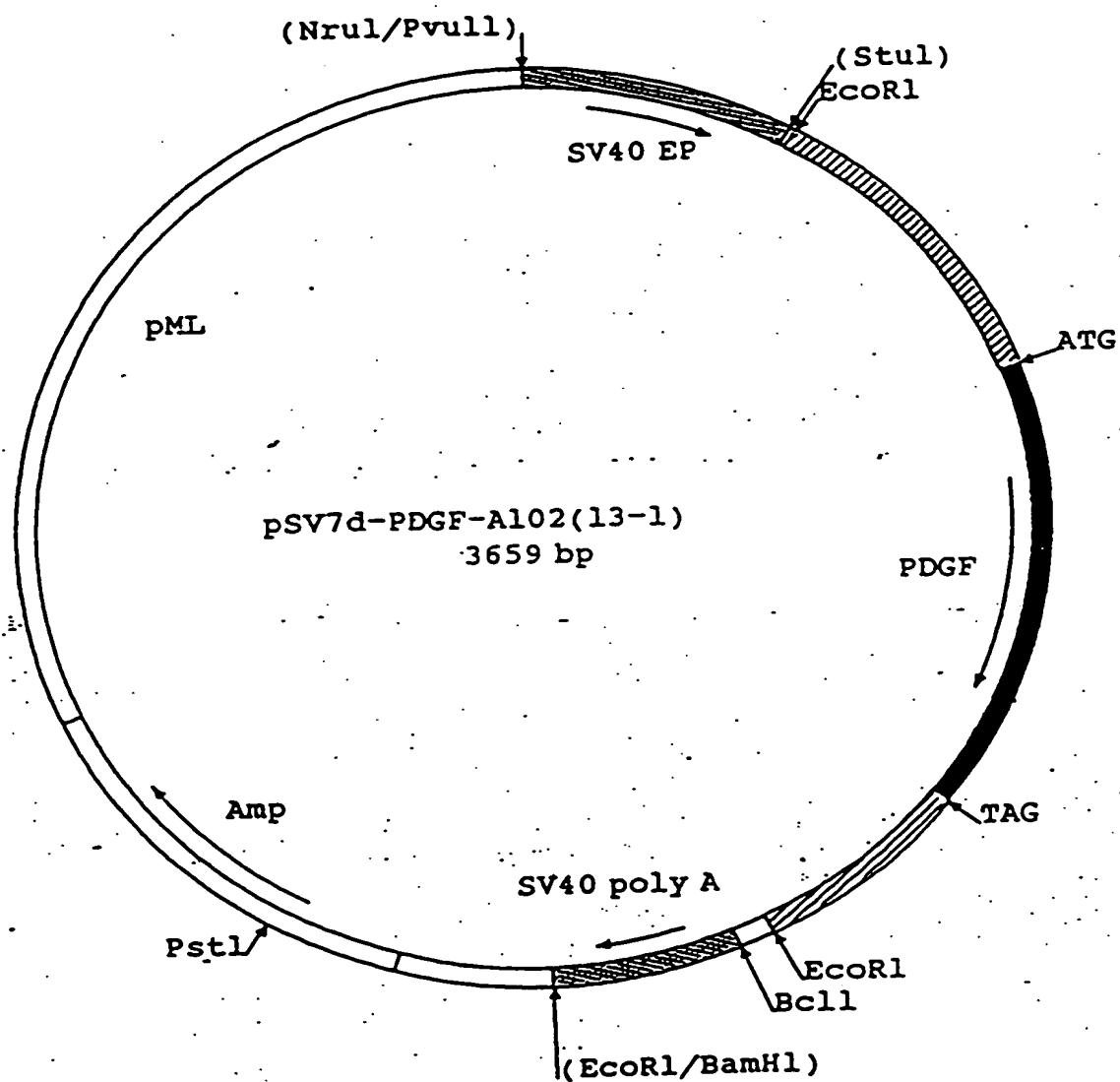
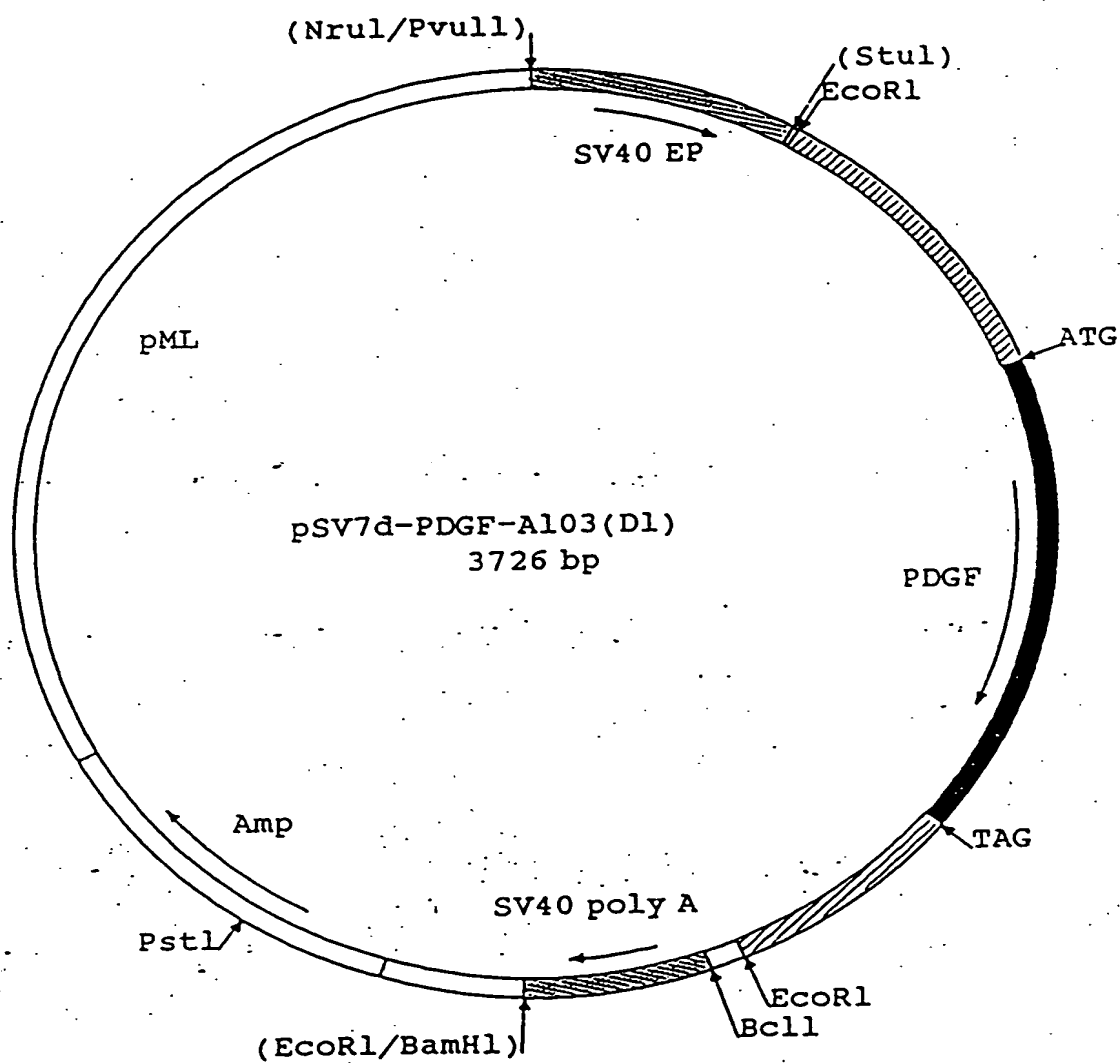
Figure 3.

Figure 4.



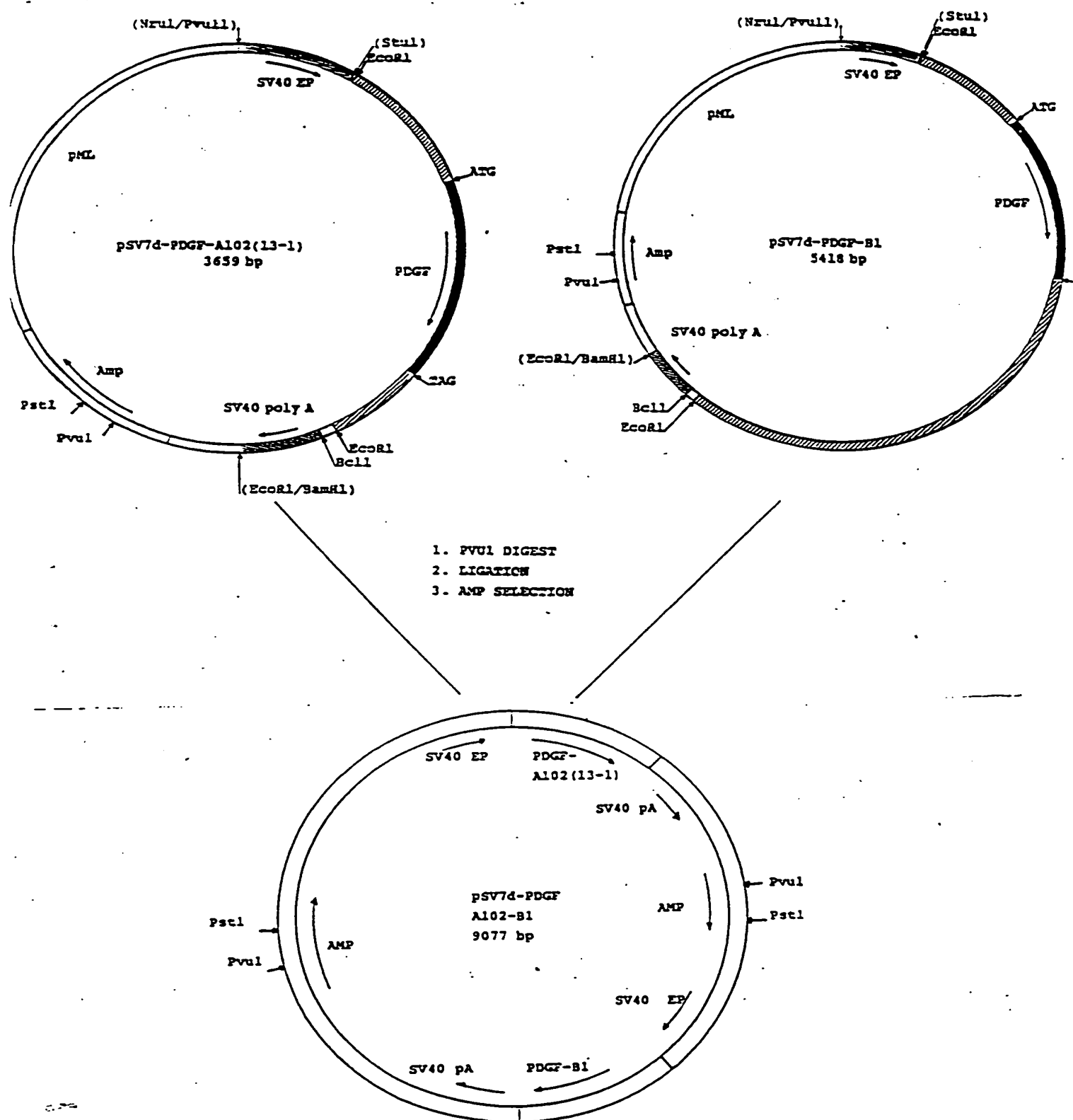


Figure 5

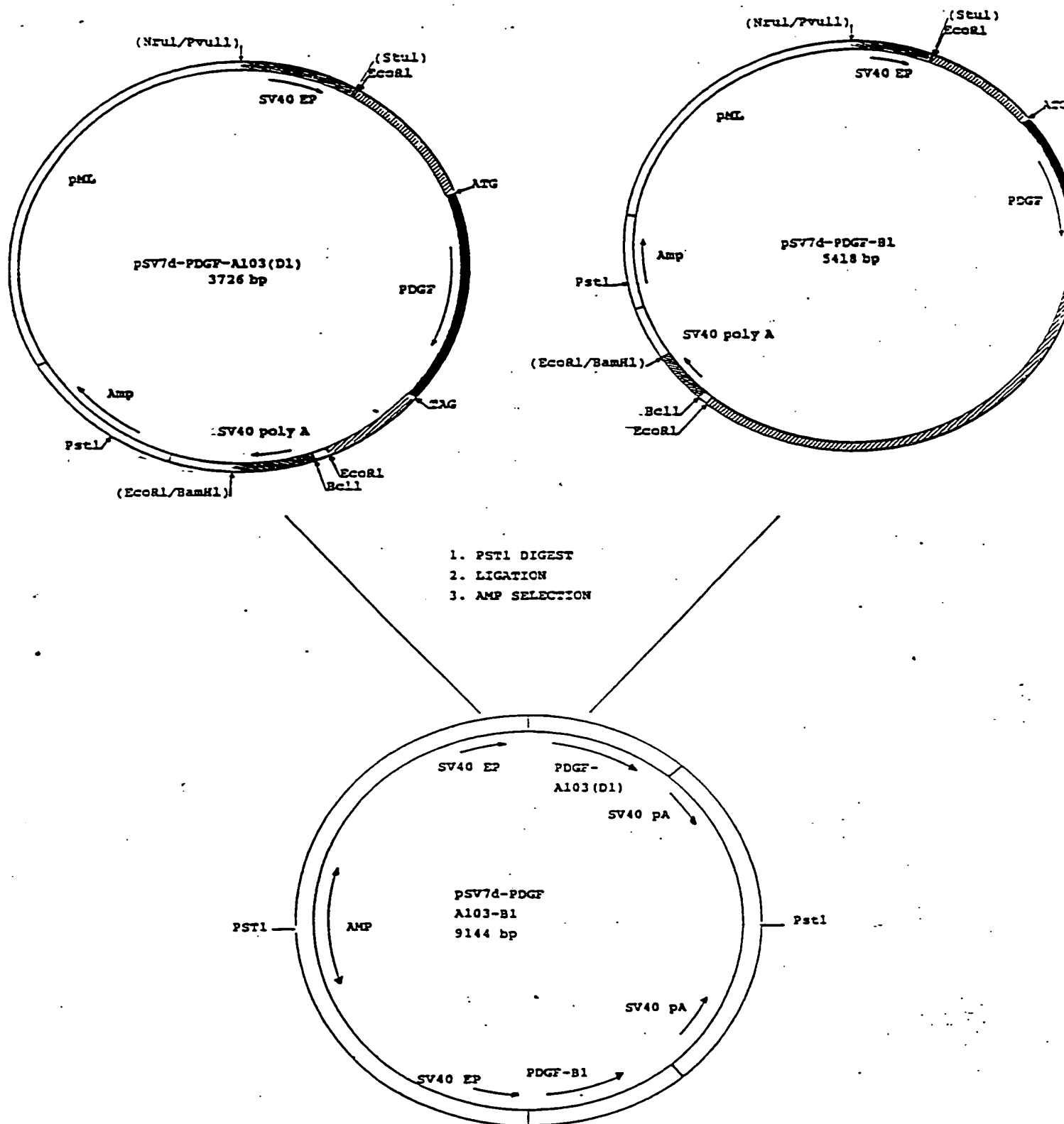


Figure 6

Sequence of SV: early promoter, polylinker, and SV40
poly A addition region from pSV7d

```

| 1/2 PVuII site SV40 early promoter --->
CTGTGGAATGTGTGTCAGTTAGGGTGTGGAAAGTCCCCAGGCTCCCCAGCAGGCAGAAGT
GACACCTTACACACAGTCAATCCACACCTTTCAGGGGTCCGAGGGGTCGTCCGTCTTCA

61  ATGCAAAGCATGCATCTCAATTAGTCAGCAAGGAAAGTCCCCAGGCTCCCCAGCAGGCAG
    TACGTTTCGTACGTAGAGTTAATCAGTCGTTCTTTCAGGGGTCCGAGGGGTCGTCCGTC

68  SPH1

121  AAGTATGCAAAGCATGCATCTCAATTAGTCAGCAACCATAGTCCCGCCCCTAACTCCGCC
    TTCATACGTTTCGTACGTAGAGTTAATCAGTCGTTGGTATCAGGGCGGGGATTGAGGCGG

132  SPH1

181  CATCCCGCCCCTAACTCCGCCAGTTCGCCCATTTCTCCGCCCCATGGCTGACTAATTTT
    GTAGGGCGGGGATTGAGGCGGGTCAAGGCGGGTAAGAGGCGGGGTACCGACTGATTAAAA

223  NCOI,

241  TTTTATTTATGCAGAGGCCGAGGCCGCCTCGGCCTCTGAGCTATTCCAGAAGTAGTGAAG
    AAAATAAATACGTCTCCGGCTCCGGCGGAGCCGGAGACTCGATAAGGTCTTCATCACTTC

263  BGLI

      | -polylinker-->                                OC  OP  OP
301  AGGCTTTTTTGGAGGAGATCGAATTCGCGGTCTAGAGGATCCGTCGACCTAGATAAGTA
    TCCGAAAAAACCTCCTCTAGCTTAAGGGCCCAGATCTCCTAGGCAGCTGGATCTATTCAT

321  ECOR1, 326  SmaI  XmaI, 332  XbaI, 338  BamHI, 344  SalI,

361  ATGATCATAATCAGCCATATCACATCTGTAGAGGTTTTACTTGCTTTAAAAAACCTCCCA
    TACTAGTATTAGTCGGTATAGTGTAGACATCTCCAAAATGAACGAAATTTTTTGGAGGGT

362  BclI, 405  DraI,

421  CACCTCCCCCTGAACCTGAAACATAAAATGAATGCAATTGTTGTTGTTAACTTGTTTATT
    GTGGAGGGGGACTTGGACTTTGTATTTTACTTACGTTAACAACAACAATTGAACAAATAA

466  HpaI

481  GCAGCTTATAATGGTTACAAATAAAGCAATAGCATCACAAATTCACAAATAAAGCATTT
    CGTCGAATATTACCAATGTTTATTTTCGTTATCGTAGTGTTTAAAGTGTTTATTTTCGTAAA

541  TTTTCACTGCATTCTAGTTGTGGTTTGTCCAAACTCATCCGCTCATGAGACAATAACCTT
    AAAAGTGACGTAAGATCAACACCAACAGGTTTGAGTAGGCGAGTACTCTGTTATTGGGA

      End of SV40---->||-pBR322 (pos. 4210)->

```

Figure 7

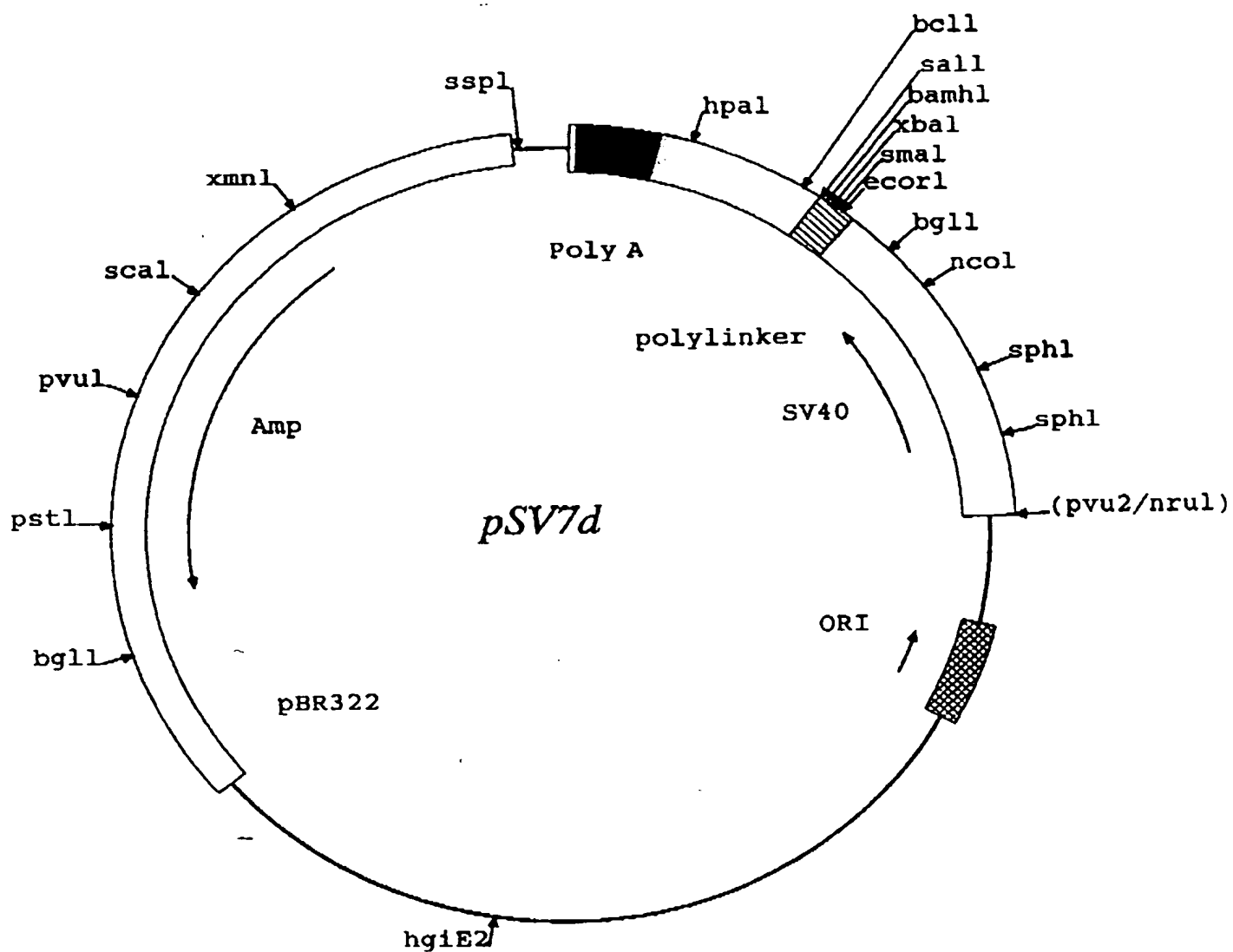


Figure 8

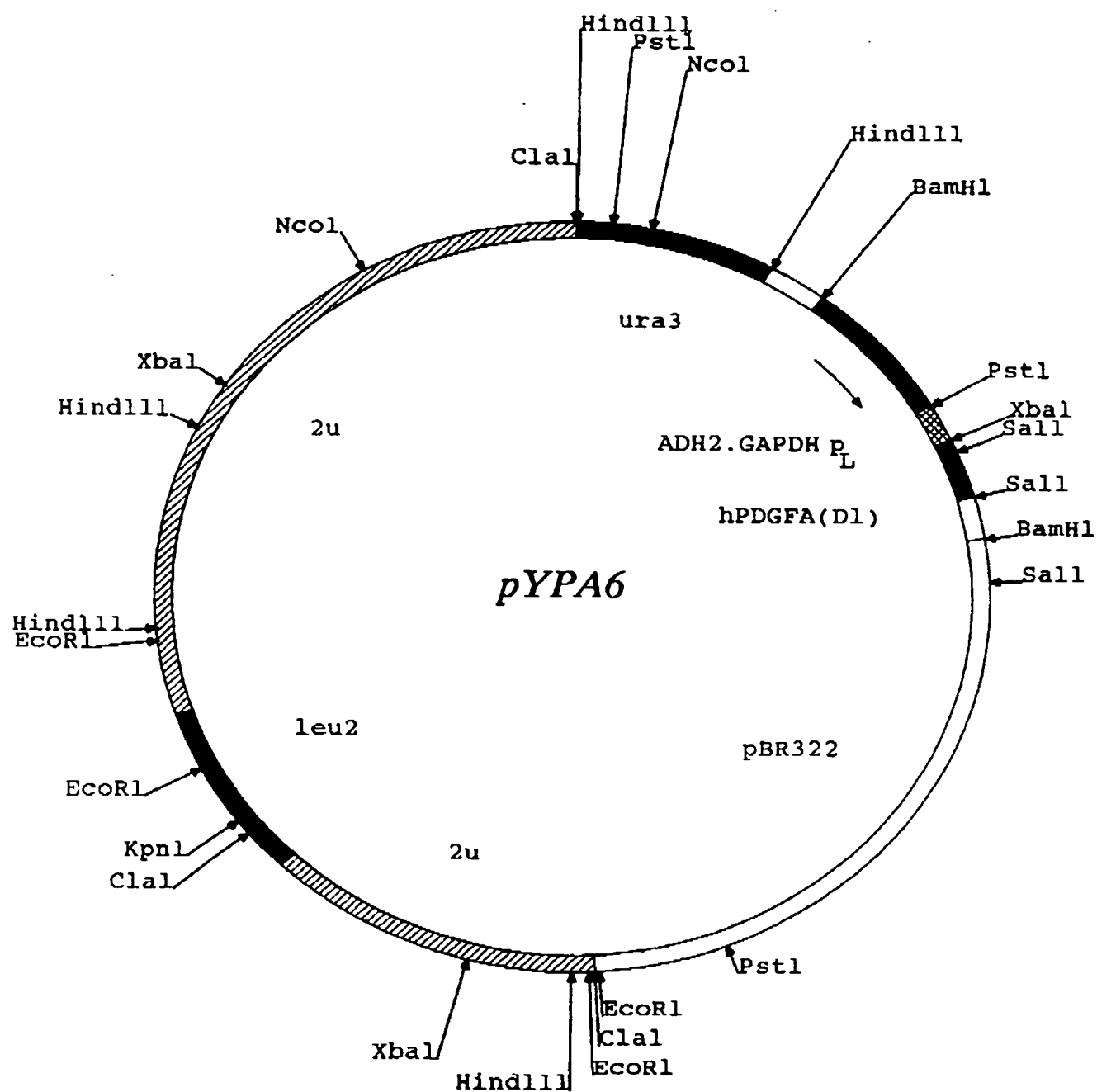


Figure 9

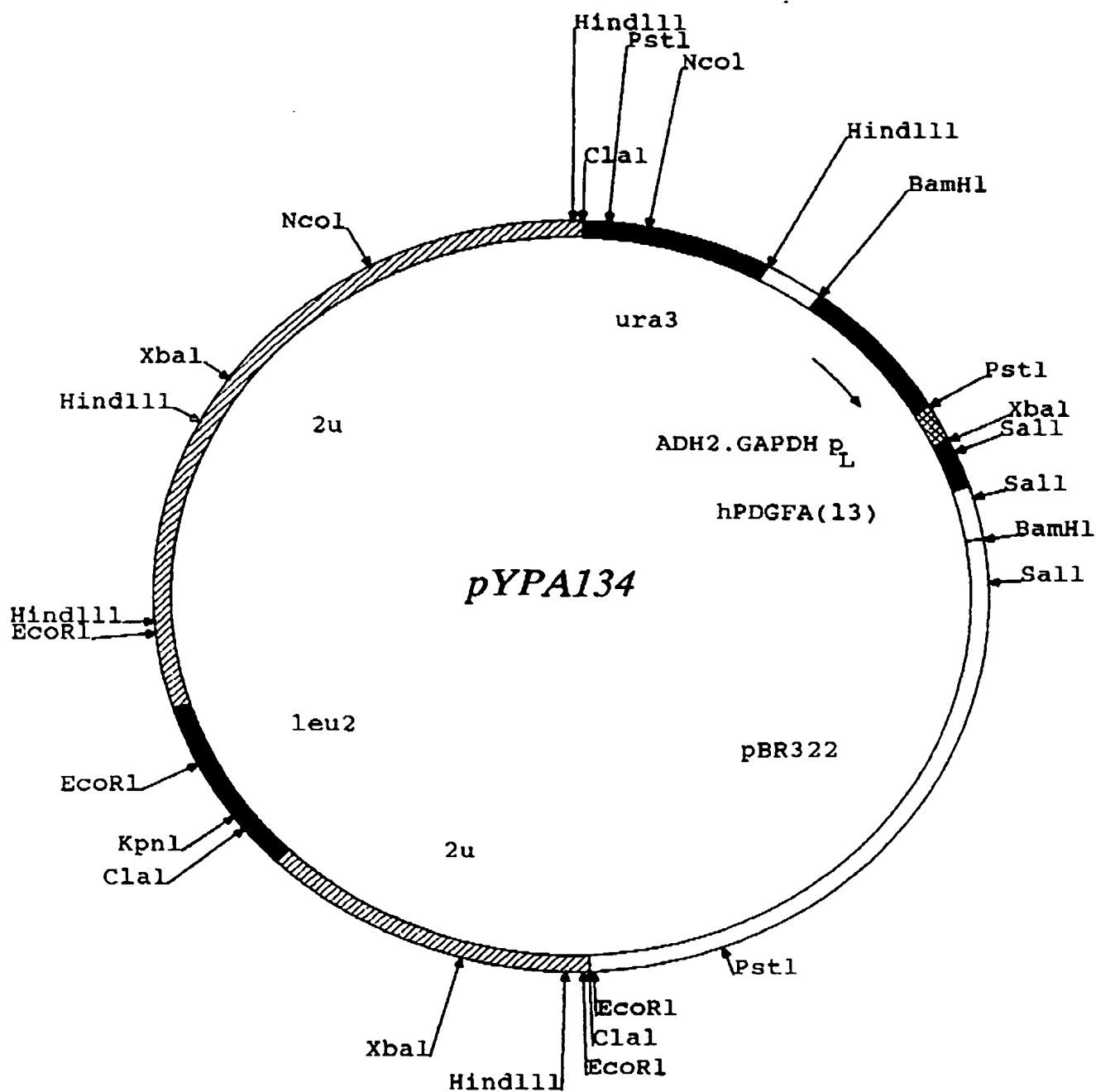


Figure 10

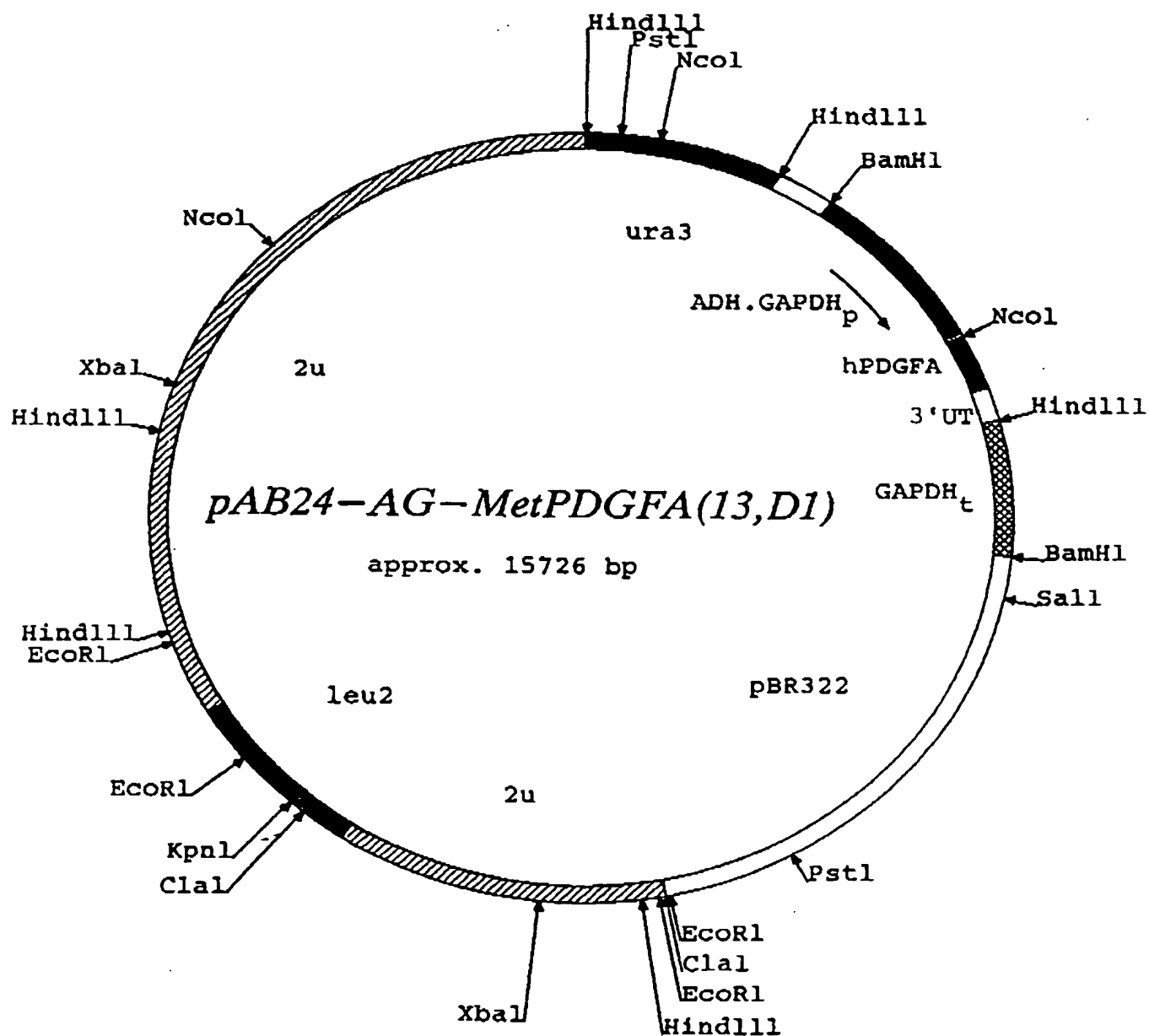


Figure 11

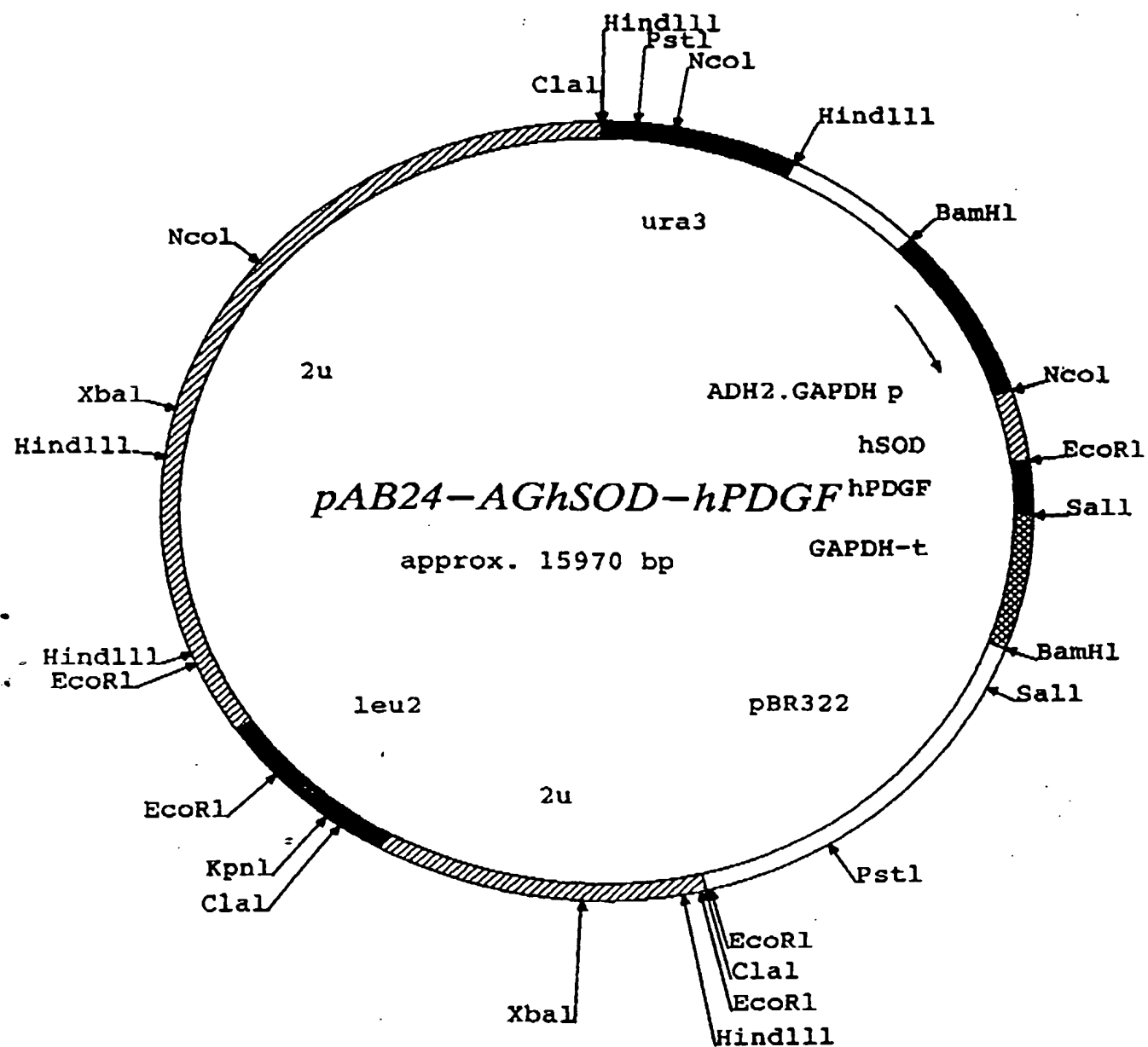


Figure 12

Figure 13.

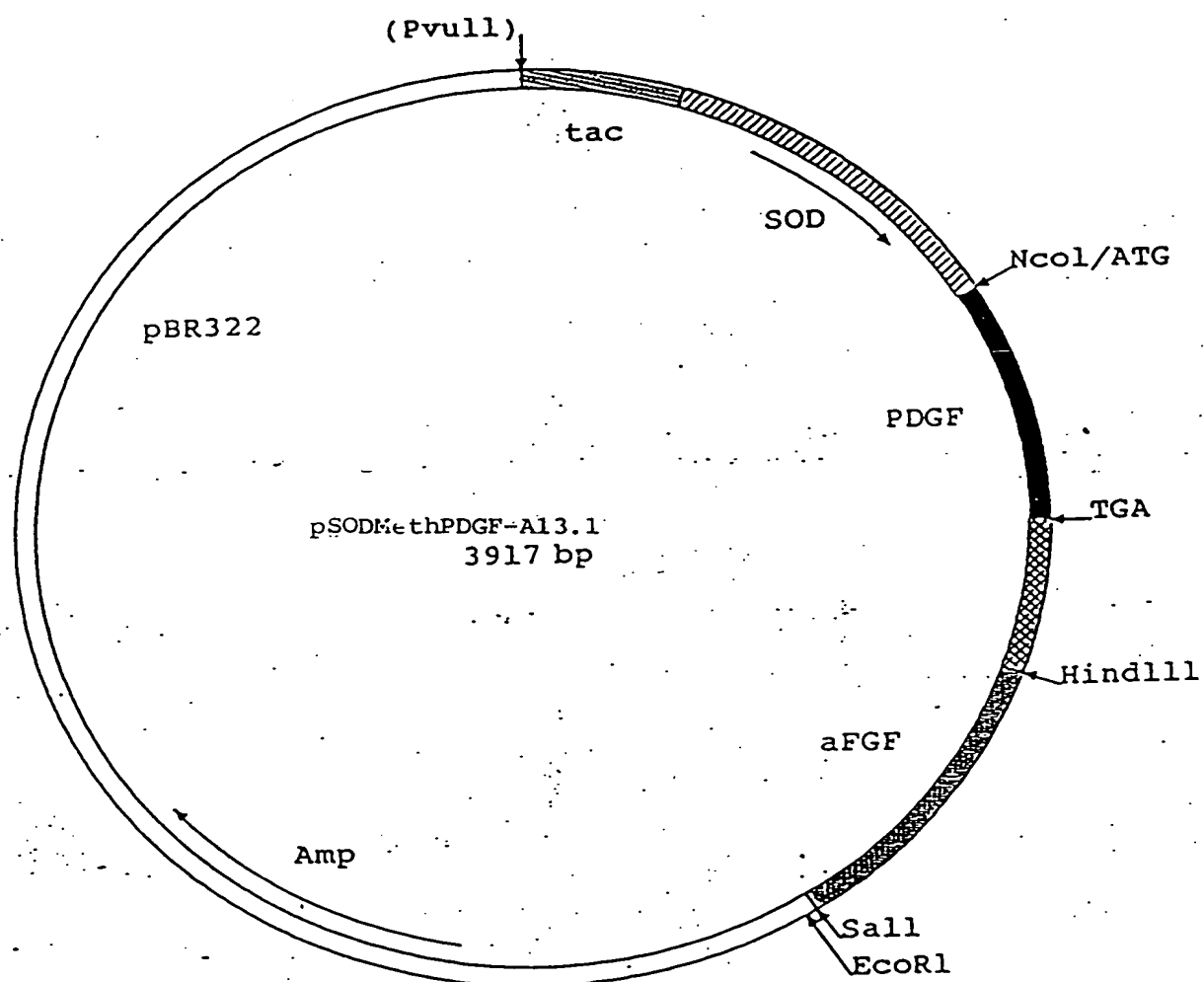


Figure 14

